

SEO AND SMW

-OR-

Why it
matters that
people find
your hard
work

■ Low Hanging Fruit

- Quick fixes for consistent search engine ranking

■ Framework and Community

- Intensive updates
- Easier to implement while building wiki

■ SMW Specific SEO

- Provided features and framework limitations

**“WE BUILT
IT AND
ALMOST
NOBODY
CAME.”**

- Dr. Mark
Greaves :
Referring to the
Linked Data
Project

OUR CURRENT SEO STRATEGY

BUILD GREAT THINGS AND LINK TO NINTENDO FRANCHISES

- Semantic Web
- WEB 3.0 – WEB 7.0
- Super Mario World
- Links on Wikipedia auto tagged ‘nofollow’
- Unreadable RDF feeds Google indexed
- Cross linking with text “smw extension”
- ???????
- PROFIT!*

* Unrealized. Estimated 2018.

There IS genuine interest in the Semantic Web and products built using the related technology stack.

COHERENT KEYWORD STRATEGY

DON'T BOTHER PAYING. NOBODY IS EVEN COMPETING

■ Zero Competition:

- Semantic Wiki
- Semantic MediaWiki
- Semantic Media Wiki
- Semantic Web Editor

■ In this listing 380,000 monthly searches are ignored with little to zero competition

Google Adwords, April 2011

<input type="checkbox"/>	Keyword	Competition
<input type="checkbox"/>	☆ semantic	<input type="text"/>
<input type="checkbox"/>	☆ semantic web technologies	<input type="text"/>
<input type="checkbox"/>	☆ what is semantic web	<input type="text"/>
<input type="checkbox"/>	☆ semantic wiki	<input type="text"/>
<input type="checkbox"/>	☆ semantic media	<input type="text"/>
<input type="checkbox"/>	☆ semantic web editor	<input type="text"/>
<input type="checkbox"/>	☆ semantic web tutorial	<input type="text"/>
<input type="checkbox"/>	☆ the semantic web	<input type="text"/>
<input type="checkbox"/>	☆ semantic media wiki	<input type="text"/>
<input type="checkbox"/>	☆ semantic web mediawiki	<input type="text"/>
<input type="checkbox"/>	☆ semantic mediawiki	<input type="text"/>
<input type="checkbox"/>	☆ linked data	<input type="text"/>
<input type="checkbox"/>	☆ semantic web	<input type="text"/>

CONSOLIDATE YOUR DOMAINS

20 URLs GOING TO THE SAME WEBSITE IS REALLY BAD

- Collect all top level domains
- Collect all sub domains
- 200 Redirect from URL to MediaWiki index.php
- 301 Redirect all remaining domains

RewriteEngine on

FIND AND REDIRECT SECONDARY DOMAINS / SUB-DOMAINS

```
rewritecond %{http_host} ^www.second-domain.com [nc]
rewriterule ^(.*)$ http://www.domain.com/$1 [r=301,nc]
```

```
rewritecond %{http_host} ^second-subdomain.domain.com [nc]
rewriterule ^(.*)$ http://www.domain.com/$1 [r=301,nc]
```

DIFFERENT ENTRANCE DEFAULTS FOR SERVER CONFIGS

```
redirect 301 /index.html /wiki/index.php
redirect 301 /index.shtml /wiki/index.php
redirect 301 /index.htm /wiki/index.php
redirect 301 /index.asp /wiki/index.php
redirect 301 /index.aspx /wiki/index.php
redirect 301 /index.cfm /wiki/index.php
redirect 301 /index.pl /wiki/index.php
redirect 301 /default.html /wiki/index.php
redirect 301 /default.htm /wiki/index.php
redirect 301 /default.asp /wiki/index.php
```

ErrorDocument 403 /notfound.html

ErrorDocument 404 /notfound.html

ErrorDocument 500 /wiki/index.php

STALE DIRECTORY OF FORMER INSTALLATION

```
redirect 301 /mediawiki/ /wiki/
```

TITLES, DESCRIPTIONS, AND HEADERS

A LITTLE APATHY HURTS A LOT OF PAGE RANK









The Rules

- Title
 - 65 characters, Unique per page
 - <site or section name> - <key words in semi-legible phrase>
- Description
 - 156 characters, unique per page, sentences can be semi-legible
- Domain /Page URL
 - 160 characters max with key words included
 - Strip session ids and non-media file extensions
- Header Tags
 - H1 & H2 : Include keywords and use within primary body blocking elements.
 - H3 – H6 : When used in conjunction with H1 & H2 their page weighting is significantly increased

TITLES, DESCRIPTIONS, AND HEADERS

A LITTLE APATHY HURTS A LOT OF PAGE RANK

■ Reasoning

Issues	Impact
The page title uses target keywords	
A single version of your website homepage is being indexed	
The page URL does not use target keywords	
The page's title tag is not unique	
The website has no accessible sitemap.xml file	
The H1 tag does not contain your target keywords	
The website provides an accessible robots.txt file	
The page contains a single H1 tag	

Provided By Optify.net, April 2011

- These elements are the top rated concerns and all major SEO services recommend correcting them first

ROBOTS, SPIDERS, AND BLACK HOLES

MEDIA WIKI BY DESIGN CAN DESTROY A SEARCH RANKING

- Due to the complexity of Media Wiki storing transactions and open editing a robots.txt and sitemap.xml are REQUIRED.
 - A wiki is technically the size of human vocabulary and 99% duplicate content create / edit pages
 - Page transactions create thousands of history logs
 - Special pages and custom functionality will provide inappropriate entrance points

These facts left unaddressed will result in heavy search engine penalties applied to your domain.

ROBOTS, SPIDERS, AND BLACK HOLES

MEDIA WIKI BY DESIGN CAN DESTROY A SEARCH RANKING

- **Example Robots.txt**
 - Bug Tracking Systems
 - Code Repositories
 - Wiki Logging
 - Wiki Editing
 - WYSIWIG panes and default create pages
 - History Logging
 - Wiki Talk Pages
 - Selected Special Pages

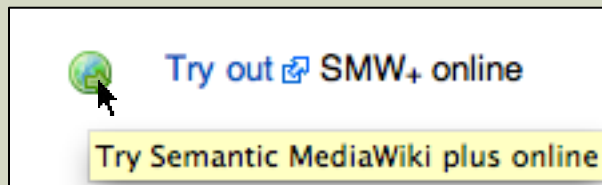
Sitemap: <http://www.url.com/sitemap.xml>

```
User-agent: *
Disallow: /smwbugs/
Disallow: /websvn/
Disallow: title=Bugzilla
Disallow: title=Special:Log
Disallow: action=annotate
Disallow: action=edit
Disallow: action=formedit
Disallow: redlink=1
Disallow: mode=wysiwyg
Disallow: action=history
Disallow: /Talk:
Disallow: title=Talk:
Disallow: /Special:Search
Disallow: /Special:Version
Disallow: title=Special:Search
Disallow: title=Special:UserLogin
```

MEDIA, ALT, AND TITLE ATTRIBUTES

THE 10% OF SEARCH RANKING EVERYBODY IGNORES

- Because early websites discovered keyword cramming and cloaking search engines started putting weight on file names, alt, and title attribute tags.
- Search engines made the mistake of assuming the Internet cares about screen readers and text browsers.
- This mistake can be exploited in a positive way by actually caring about screen readers and text browsers.
 - File names should describe the contents – Name before upload
 - Alt tags should be the normalized file name
 - Title tags should describe the section the file is within



CREATING ANCHORS WITH KEYWORDS

WHERE WAS THE LAST STREET SIGN READING ‘GO HERE’?

- Anchors on your wiki between sections should be 2-3 word descriptions of that section
- Pages will be ranked higher if keywords used in anchors match their titles
 - When creating new wiki articles take into account the article name becomes the title of the page

Example Anchor Text	Page Title
“Download Semantic MediaWiki”	SMW : Download Semantic MediaWiki
“SMW Community Forum”	SMW : Community and Development Forum
“Purchase SMW+ License”	SMW+ : Purchase Semantic MediaWiki Plus
“Business Semantic MediaWiki”	SMW : Small Business and Academic Portal

CONSISTENT IN LINKS TO YOUR WIKI

YOUR E-REPUTATION MATTERS TO SEARCH ENGINES

- Same rules apply to in links as internal anchors between pages and sections
- Links with the attribute rel="nofollow" are not followed or applied to your wikis search engine ranking.
 - Wikipedia.com
 - Lesser known search portals
 - News sites with millions of daily readers
- Google Page Rank heavily driven by quality in links to your site
 - <http://www.adsensetricks.org/PRCalcChart.htm>

Buying in links is risky! Many agencies selling a number of links for a price are dumping them on low ranked blogs who cloak the links.

CONSISTENT OUT LINKS FROM YOUR WIKI

NOT EVERY WEBSITE SHOULD BE TREATED THE SAME

- Your page ranking and search engine results are also effected by the sites you link and HOW they are linked
 - Over 6 months old with page rank of zero remove the link
 - Large corporations don't need your page rank status. Use rel="nofollow" for their pages
 - Most links on your wiki will have an automated rel="nofollow" added to them
 - There are custom solutions for removing the nofollow attributes that should be used when linking other SMW sites and MediaWiki resources.
 - Link to sites how you want linked from sites. If you don't know which keywords a wiki is targeting use terms from the page title

IMPROVE PAGES WITH SEMANTIC TAGS

SEMANTIC MARKUP OF PAGES IS THE FUTURE OF SEARCH ENGINE RANKING

- Marking up existing pages can take the form of Microformats, XML attributes, or RDFa
- Example of Amazon's current markup for search results :


```
<div>  
<a href="http://www.amazon.com/gp/pdp/profile/A2G8PQ9HNUY6NA/">  
<span property="v:reviewer" about="profile/A2G8PQ9HNUY6NA/">Marian the Librarian</span>  
</a> (NY, NY) - <span property="v:dtreviewed">1st April 2009</span>  
</div>
```

- Popular markup formats include:
 - RDFa
 - Google Rich Snippets
 - Good Relations
 - Yahoo! Search Monkey
 - hCalendar, hCard, hReview
- Not actually proven to render consistent results across websites

TRACK AND IMPROVE IMPORTANT PAGES

YOU CAN'T ALWAYS CHOOSE WHICH PAGE IS RANKED HIGHLY

- An example of unexpected pages with high ranking:

[Help:Semantic MediaWiki User Manual - SMW+ A Semantic Web ...](#) 
Feb 21, 2011 ... user manual article for **Semantic MediaWiki 1.5.2** ... **Semantic MediaWiki** (SMW) is a free extension of MediaWiki – the wiki-system powering ...
[smwforum.ontoprise.com/.../Help:Semantic_MediaWiki_User_Manual](#) - [Cached](#) - [Similar](#)

- This is a help manual page that is listed on the first page for a keyword term
 - Current content should not be modified but extended
 - Out links from this page should be removed or kept few and relevant
 - Links to internal targeted sections should be added and made visible on this page
 - Semantic markup should be added to the page body and refined for the keyword giving the page a high rank

DON'T CHEAT THE SEARCH ENGINES

WITH RISK COMES REWARD... UNTIL YOU ARE BANNED

■ Link Farming

- Buy a dozen domains and skin off one framework
- All domains link to your new website in slightly different ways cramming your keywords

■ Paying the Link Farmers

- No. They cannot get you linked on 500, Page Rank 4+, websites for \$99 without owning the Link Farm or degrading the Page Rank of naïve blog owners.

■ Cloaking

- Display one thing to the SE Spider and something else to users
- EVEN FOR MOBILE SPIDERS DO NOT DO THIS

■ Off Page Linking and Spamming

- Search engines understand if your div is positioned off browser or display is set to none
- If you must do this for menus use HTML5 and specify <nav>, or in lieu of that use an ID or Class name of “menu”, “nav”, or “navigation”

LIMITATIONS OF SEO ON MEDIA WIKI

IT WAS CREATED TO BE FUNCTIONAL, NOT ALWAYS POPULAR

- Customization of page titles and descriptions is time consuming and requires custom extensions
 - Pages titles are bound to the article title, namespace, and URL
- No way by default to define meta descriptions and keyword listings within the article
- Nofollow attributes are either always on or always off, and require specialty code to remove them from a single anchor
- Most of the wiki can be classified as duplicate content and requires extensive robots.txt and sitemap.xml files
- Irregular entrance points as certain articles gain in popularity with no consideration for the entire topic of the wiki
- Easy to spam

SEARCH ENGINE WEBMASTER ACCOUNTS

ANALYTICS TRACK. THESE ACCOUNTS EXPEDITE SUBMISSION

- Google Webmaster Tools
 - <http://www.google.com/webmasters/tools/>
- Yahoo! Webmaster Tools
 - <http://siteexplorer.search.yahoo.com/>
- Bing Webmaster Tools
 - <http://www.bing.com/toolbox/webmasters/>
- DMOZ Open Directory
 - <http://www.dmoz.org/docs/en/add.html>
- Bulk Sitemap Submission (Yahoo!, Google, Bing, Ask)
 - <http://www.sitemapwriter.com/notify.php>
- Region Specific Search Engines
 - Yandex – Russia
 - Baidu – China
 - Full listing : <http://www.searchenginecolossus.com/>

SEO TOOLS AND WIKI SUGGESTIONS

INFORMATION, TOOLS, AND WISH LISTS

- <http://www.seowarrior.net/> - SEO Warrior, O'Reilly
 - Updated versions of PERL scripts for site wide statistics compilation
 - Book usually updated every 18 months, site keeps e-version
- Analytics Tracking
 - 1 or 2 Online Accounts (Google Analytics)
 - 1 to 3 Server Side Systems (Splunk SEO App, Pwik, Awstats)
- MediaWiki Extensions
 - http://www.mediawiki.org/wiki/Extension:Advanced_Meta
 - http://www.mediawiki.org/wiki/Extension:Add_HTML_Meta_and_Title
 - Somebody create an SEO centric skin or update existing skins?
 - Administration tools for bulk renaming files and pages complete with 301 Redirect tags created
- SEO Tracking Pay Services
 - <http://www.seomoz.org/>
 - <http://www.optify.net>

QUESTIONS?

Comments,
Requests,
Ideas,
Complaints,
Would you
like to know
more?